DOCUMENT RESUME

ED 353 296 TM 019 311

AUTHOR Paul, Richard; Nosich, Gerald M.
TITLE A Model for the National Assessment of Higher Order Thinking.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 92
NOTE 26p.
AVAILABLE FROM Foundation for Critical Thinking, 4655 Sonoma Mountain Road, Santa Rosa, CA 95404 ($3).
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Academic Standards; *Cognitive Measurement; *Critical Thinking; Educational Assessment; Educational Change; Elementary Secondary Education; *Evaluation Criteria; Intelligence; Models; National Competency Tests; National Programs; Problem Solving; *Student Evaluation; Testing Programs; *Thinking Skills
IDENTIFIERS *Elements of Reasoning; Reform Efforts

ABSTRACT

This document provides a model for the national assessment of higher order thinking and consists of a preface and four main sections. The preface discusses the problem of lower order learning, summarizes the state of research into critical thinking and instructional reform, and explains the structure of the paper. Section 1 of the paper describes 21 criteria for higher order thinking assessment. Section 2 makes the case for how "a rich substantive concept of critical thinking" meets these 21 criteria, and highlights the dangers of a non-substantive concept of critical thinking. Section 3 outlines four domains of critical thinking (elements of thought, abilities or basic modes of reasoning, affective dimensions, and universal intellectual standards). Section 4 makes substantive recommendations regarding how to assess the various domains of critical thinking, the test strategies that may be used (including multiple rating items and essay items), the value of the proposed strategy for educational reform, and the suggested implementation of the proposed assessment. An assessment of the results of critical thinking instruction should focus on thinking within the framework of particular academic subjects and thinking in interdisciplinary contexts that are important for functioning as an autonomous well-informed productive member of a democracy. Seven figures are included. (RLC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

A Model for the National Assessment of Higher Order Thinking

ED353296

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

RENEE DENISE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

with Gerald M. Nosich

Abstract

This paper, co-authored by Richard Paul and Gerald Nosich, was commissioned by the United States Department of Education, Office of Educational Research and Improvement of the National Center for Education Statistics. It provides exactly what its title implies: a model for the national assessment of higher order thinking. The paper consists of a preface and five main sections.

The preface delineates the problem of lower order learning, summarizes the state of research into critical thinking and educational reform, and explains the five-part structure of the paper. The first main section of the paper states and explicates 21 criteria for higher order thinking assessment. The second section makes the case for how a "rich, substantive concept of critical thinking" meets those criteria. In making this case, Paul and Nosich spell out the dangers of a non-substantive concept of critical thinking. The third section of the paper spells out four domains of critical thinking: elements of thought, abilities, affective dimensions, and intellectual standards. The fourth section of the paper makes substantive recommendations regarding how to assess the various domains of critical thinking, the test strategies that may be used, the value of the proposed strategy for the reform of education, and the suggested implementation of the proposal.

✦ *Preface:*
The Problem of Lower Order Learning

Virtually all informed commentators agree that schooling today does not foster the "higher order thinking skills and abilities which represent the "basics" of the future. America 2000, President Bush's education initiative, seeks to bring schooling in line with changing global and economic conditions, to engender sweeping educational reform in what are now admittedly largely static institutions, systems highly resistant to substantial change. America 2000 raises the following vital question: "How can we reverse the pervasive emphasis in education on lower rather than on higher order learning, on recall rather than on reasoning, on students merely 'reproducing' rather than 'producing' knowledge?"

The state of research regarding this problem was summarized recently by Mary Kennedy in an article for the Kappan:

> ...national assessments in virtually every subject indicate that, although our students can perform basic skills pretty well, they are not doing well on thinking and reasoning. American students can compute, but they cannot reason.... They can write complete and correct sentences, but they cannot prepare arguments.... Moreover, in international comparisons, American students are falling behind ... particularly in those areas that require higher order thinking.... Our students are not doing well at thinking, reasoning, analyzing, predicting, estimating, or problem solving.

In this summary, Dr. Kennedy linked the problem to the established mode of instruction:

> ...teachers are highly likely to teach in the way they themselves were taught. If your elementary teacher presented mathematics to you as a set of procedural rules with no substantive rationale, then you are likely to think that this is what mathematics is and that this is how mathematics should be studied. And you are likely to teach it in this way. If you studied writing as a set of grammatical rules rather than as a way to organize your thoughts and to communicate ideas to others, then this is what you will think writing is, and you will probably teach it so.... By the time we complete our undergraduate education, we have observed teachers for up to 3,060 days.

Though not as commonly realized, this problem of the dominance of lower order learning is as serious in post-secondary as it is in primary and secondary education. In both undergraduate and graduate programs students are typically enrolled in content heavy courses taught by professors who feel a greater obligation to cover subject matter through lecture than to generate thought-provoking activities or assignments that may seriously reduce what they can cover or significantly add to their work load, or both.

Alan Schoenfeld has explored this problem with respect to both pre-secondary and post-secondary mathematics instruction. To illustrate the detailed nature of what Schoenfeld's research is disclosing, here is a summary from one of his studies:

BEST COPY AVAILABLE

> At the University of Rochester 85% of the freshman class takes calculus, and many go on ... [but] most of these students will never apply calculus in any meaningful way (if at all) in their studies, or in their lives. They complete their studies with the impression that they know some very sophisticated and high-powered mathematics. They can find the maxima of complicated functions, determine exponential decay, compute the volumes of surfaces of revolution, and so on. But the fact is that these students know barely anything at all. The only reason they can perform with any degree of competency on their final exams is that the problems on the exams are nearly carbon copies of problems they have seen before; the students are not being asked to think, but merely to apply well-rehearsed schemata for specific kinds of tasks. Tim Keiter and I studied students' abilities to deal with pre-calculus versions of elementary word problems.... We were not surprised to discover that only 19 of 120 attempts at such problems ... yielded correct answers, or that only 65 attempts produced answers of any kind.

Schoenfeld summarizes the results, in general, of research into mathematics instruction as follows:

> In sum: all too often we focus on a narrow collection of well-defined tasks and train students to execute those tasks in a routine, if not algorithmic fashion. Then we test the students on tasks that are very close to the ones they have been taught. If they succeed on those problems, we and they congratulate each other on the fact that they have learned some powerful mathematical techniques. In fact, they may be able to use such techniques mechanically while lacking some rudimentary thinking skills. To allow them, and ourselves, to believe that they 'understand' the mathematics is deceptive and fraudulent.

There is good reason, in our view, to link instructional reform with the need for a special emphasis on critical thinking, problem solving, and communication skills, for it is precisely these higher order thinking skills that are routinely sacrificed when coverage and lower order recall dominate the classroom at either the pre- or post-secondary level, as they now do.

✦ *The State of Research into Critical Thinking and Instructional Reform*

One major value of the last ten years of research into critical thinking is the focus on the need for reform of instruction at all levels: on the need for students to reason mathematically in mathematics courses, to reason historically in history courses, to reason scientifically in science courses, to reason sociologically in sociology courses. Indeed, critical thinking research has emphasized three basic needs for all learning: for all students to reason out all basic concepts and understandings, to reason to all basic conclusions and solutions, and to reason through and across the curriculum.

This emphasis has been embedded in the structure of the 11 major international conferences on research into critical thinking and educational reform (1980–1991) held at Sonoma State University. The 1991 Conference attracted 1400 registrants from 20 countries and featured over 300 sessions representative of education from kindergarten through graduate school. This same emphasis is reflected in the 25 or so other conferences focused on critical thinking in the last ten years (at Harvard, the University of Chicago, Montclair State, Oakton College, and elsewhere) and in most of the articles published concerning critical thinking.

What is more, the research into critical thinking has focused not only on the cultivation of reasoning in all subjects and at all educational levels, but also on generalizable standards for the assessment of reasoning as well. The concepts and distinctions embedded in critical thinking research are, as a result, well-suited for the design of a process to assess higher order thinking. In this paper, we shall set out both the conceptual foundations for such a process as well as a viable model for carrying out that process.

Before we spell out the detailed structure of this paper, however, it is important to note that the concept of critical thinking has not played a central role in the design of educational assessment instruments to date, principally because the concept has been developed extensively only over the last ten years, and therefore has not had time to permeate already developed assessment tools. Now that we possess a rich, substantive concept, however, we have an unprecedented opportunity to assess central rather than peripheral aspects of critical thinking, and to do so in an authentic and representative way. If anything less than this concept and its central aspects is assessed, the ultimate goal of fostering higher order thinking as an academic, social, and vocational need will be ill served.

✦ *The Structure of the Paper*

The substance of this paper is divided into four sections, each focused on a major question, as follows:
Section One: What should be the main objectives of a process to assess higher order thinking?
Section Two: How does a rich, substantive concept of critical thinking meet these criteria?

a) What is included in a rich, substantive concept of critical thinking?

b) How, specifically, does this concept meet the criteria?

c) What, specifically, are the dangers of a non-substantive concept of critical thinking?

Section Three: What are the four component domains of critical thinking and the implications of each of these domains for the assessment of higher order thinking?
Section Four: What is the most workable solution to the design of a process to assess higher order thinking, given the findings in the three sections above?

BEST COPY AVAILABLE

The first section of the paper formulates 21 objectives that should be met by any process adequate to the task. The second outlines the basic concept of critical thinking which informs the paper and explains how a rich, substantive concept of critical thinking, grounded in the research on critical thinking, provides a plausible foundation for accomplishing these objectives. The third section of the paper explicates the four domains essential to critical thinking:

A) The *Elements of Thought* (eight essential dimensions of all reasoning crucial for understanding and assessing reasoning),

B) *Abilities* (basic modes of reasoning — including reading, writing, speaking, and listening — that represent modal "orchestrations" of the elements of thought),

C) *Traits of Mind* (the affective dimensions without which critical thinking skills are merely episodically used, and often in a limiting rather than an expansive manner), and

D) *Universal Intellectual Standards* (presupposed by critical thinking).

As we give a brief explication of the elements of thought, the abilities, essential traits of mind, and intellectual standards, we briefly comment on the implications for assessment purposes of each conception.

In the fourth and final section of the paper, we lay out our recommendations for a process and a time-table for assessing higher order thinking.

✦ *Section One: Objectives*

What should be the main objectives of a process to assess higher order thinking?

1) It should assess students' skills and abilities in analyzing, synthesizing, applying, and evaluating information.

2) It should concentrate on thinking skills that can be employed with maximum flexibility, in a wide variety of subjects, situations, contexts, and educational levels.

3) It should account for both the important differences among subjects and the skills, processes, and affective dispositions that are crucial to all the subjects.

4) It should focus on fundamental, enduring forms of intellectual ability that are both fitted to the accelerating pace of change and deeply embedded in the history of the advancement of the disciplines.

5) It should readily lead to the improvement of instruction.

6) It should make clear the inter-connectedness of our knowledge and abilities, and why expertise in one area cannot be divorced either from findings in other areas or from a sensitivity to the need for interdisciplinary integration.

7) It should assess those versatile and fundamental skills that are essential to being a responsible, decision-making member of the work-place.

8) It should be based on clear concepts and have well-thought-out, rationally articulated goals, criteria, and standards.

9) It should account for the integration of communication skills, problem-solving, and critical thinking, and it should assess all of them without compromising essential features of any of them.

10) It should respect cultural diversity by focusing on the common-core skills, abilities, and traits useful in all cultures.

11) It should test for thinking that is empowering and that, when incorporated into instruction, promotes (to quote the September, 1991 Kappan) "the active engagement of students in constructing their own knowledge and understanding."

12) It should concentrate on assessing the fundamental cognitive structures of communication, for example:

with reading and listening, the ability to

- create an accurate interpretation,
- assess the author's or speaker's purpose,
- accurately identify the question-at-issue or problem being discussed,
- accurately identify basic concepts at the heart of what is said or written,
- see significant implications of the advocated position,
- identify, understand, and evaluate the assumptions underlying someone's position,
- recognize evidence, argument, inference (or their lack) in oral and written presentations,
- reasonably assess the credibility of an author or speaker,
- accurately grasp the point of view of the author or speaker,
- empathetically reason within the point of view of the author or speaker.

with writing and speaking, the ability to

- identify and explicate one's own point of view and its implications,
- be clear about and communicate clearly, in either spoken or written form, the problem one is addressing,
- be clear about what one is assuming, presupposing, or taking for granted,
- present one's position precisely, accurately, completely, and give relevant, logical, and fair arguments for it,
- cite relevant evidence and experiences to support one's position,
- see, formulate, and take account of alternative positions and opposing points of view, recognizing and evaluating evidence and key assumptions on both sides,
- illustrate one's central concepts with significant examples and show how they apply in real situations,
- empathetically entertain strong objections from points of view other than one's own.

13) It should assess the skills, abilities, and attitudes that are central to making sound decisions and acting on them in the context of learning to understand our rights and responsibilities as citizens, as well-informed and thinking consumers, and as participants in a symbiotic world economy.

14) It should avoid any reductionism that allows a multi-faceted, theoretically complex, and authentically usable body of abilities and dispositions to be assessed by means of oversimplified parts that do not adequately reflect the whole.

15) It should enable educators to see what kinds of skills are basic for the future.

16) It should be of a kind that will assess valuable skills applied to genuine problems as seen by a large body of the populace, both inside and outside of the educational community.

17) It should include items that assess both the skills of thoughtfully choosing the most reasonable answer to a problem from among a pre-selected set and the skills of formulating the problem itself and of making the initial selection of relevant alternatives.

18) It should contain items that, as much as possible, are examples of the real-life problems and issues that people will have to think out and act upon.

19) It should be affordable.

20) It should enable school districts and educators to assess the gains they are making in teaching higher order thinking.

21) It should provide for a measure of achievement against national standards.

✦ *Section Two: Critical Thinking and Criteria for Assessment*

✦ *What Is Included in a Rich, Substantive Concept of Critical Thinking?*

Most of the language we shall use is drawn from draft statements of the *National Council for Excellence in Critical Thinking Instruction.* The National Council has been established precisely to articulate standards in critical thinking by 50 key leaders in critical thinking research and 105 leading educators. It is in the process of establishing regional offices and setting up 75 research-based committees to articulate the state of research in the field.

NATIONAL COUNCIL DEFINITION

Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action.

This is the working definition of the National Council for Excellence in Critical Thinking Instruction. Though the definition as well as the other draft statements of the Council are subject to modification and refinement, the basic idea is one that is common to practitioners and researchers in critical thinking.

GLOSS ON THE DEFINITION

"In its exemplary form, [critical thinking] is based on universal intellectual values that transcend subject-matter divisions: clarity, accuracy, precision, consistency, relevance, sound evidence, good reasons, depth, breadth, and fairness." (National Council Draft Statement)

a) "It entails the examination of those structures or elements of thought implicit in all reasoning: purpose; problem, or question-at-issue; assumptions; concepts; empirical grounding; inferences; implications and consequences; objections from alternative viewpoints, and frame of reference." (National Council Draft Statement)

b) It entails larger-scale abilities of integrating elementary skills in such a way as to be able to apply, synthesize, analyze, and evaluate complicated and multidimensional issues. These include such abilities as clarifying issues, transferring insights into new contexts, analyzing arguments, questioning deeply, developing criteria for evaluation, assessing solutions, refining generalizations, and evaluating the credibility of sources of information. Among the abilities are included also the central forms of communication: critical

reading, writing, speaking, and listening. Each of them is a large-scaled mode of thinking which is successful to the extent that it is informed, disciplined, and guided by critical thought and reflection. (Paraphrased from National Council Draft Statement.)

c) Critical thinking entails the possession and active use of a set of traits of mind and affective dimensions: independence of thought, fairmindedness, intellectual humility, intellectual courage, intellectual perseverance, intellectual integrity, curiosity, confidence in reason, and the willingness to see objections, to enter sympathetically into another's point of view, and to recognize one's own egocentricity or ethnocentricity. (Paraphrased from National Council Draft Statement.)

Critical thinking — in being responsive to variable subject areas, issues, and purposes — is incorporated in a family of interrelated modes of thinking, among them: scientific thinking, mathematical thinking, historical thinking, anthropological thinking, economic thinking, moral thinking, and philosophical thinking (National Council Draft Statement).

✦ *How Does a Rich, Substantive Concept of Critical Thinking Meet the 21 Criteria?*

In our view, a rich, substantive concept of critical thinking, and it alone, provides an intelligible and workable means of meeting all 21 criteria. In this section we will briefly consider each objective in turn, not as a definitive response to the criteria, but merely to suggest the fuller response in Section Three below.

CRITERION # 1

Can it be used to test information processing skills? Critical thinking includes at its core "a set of information and belief generating and processing skills and abilities."

CRITERION # 2

Can it be used to test flexible skills and abilities that can be used in a wide variety of subjects, situations, contexts, and educational levels? Since the art of critical thinking "entails proficiency in the examination of those structures or elements of thought implicit in all reasoning — purpose, problem or question-at-issue, assumptions, concepts, empirical grounding, reasoning leading to conclusions, implications and consequences, objections from alternative viewpoints and frames of reference" — it provides for maximum flexibility of use. It can be used in any subject, with respect to any situation to be figured out, in any context in which reasoning is germane, and, if adapted to the proficiency of students, at any educational level.

CRITERION # 3

Can it account for important differences among the subject areas? Subjects differ not because some make assumptions and others do not, not because some pose questions or problems and others do not, not because some have purposes and others do not, but rather because each has somewhat different purposes, and hence asks somewhat different questions, poses somewhat different problems, gathers somewhat different evidence, uses somewhat different concepts, etc. Critical thinking highlights these differences while underlining common structural features.

CRITERION # 4

Can it be used to focus on fundamental abilities fitted to the accelerating pace of change and embedded in intellectual history? Basic critical thinking skills and abilities are readily shown to be implicit in the rational development and critique of ideas at the core of intellectual history. They explain, for example, how new disciplines emerge from established ones: that is, by asking new questions, pursuing new purposes, framing new concepts, gathering new data, making new assumptions, reasoning in new directions, etc. They also explain how it is that a new field of study can ground itself, even at the outset, on definite intellectual standards that transcend any particular academic field: clarity, precision, accuracy, relevance, consistency, evidentiary force, valid reasoning, consistency . . . (standards implicit in the history of critical thinking and rational discourse in every domain).

CRITERION # 5

Can it be used to improve instruction? Critical thinking is not an isolated good, unrelated to other important goals in education. Rather it is a seminal goal which, done well, simultaneously facilitates a rainbow of other ends. It is best conceived, therefore, as the hub around which all other educational ends cluster. For example, as students learn to think more critically, they become more effective readers, writers, speakers, and listeners because each ability requires well-reasoned thought. They increase their mastery of content because all content is embedded in a system of understandings which, to be grasped, must be reasoned through. They become more proficient in — because they must be practiced within — a variety of modes of thinking: for example, historical, scientific, and mathematical thinking. Self-confidence increases with the intellectual empowerment critical thinking engenders. Finally, they develop skills, abilities, and traits of mind (intellectual discipline, intellectual perseverance, intellectual humility, intellectual empathy, intellectual integrity, ...) crucial to success in the educational, professional, and everyday world.

CRITERION # 6

Can it make clear the inter-connectedness of our knowledge and abilities, and why expertise in one area cannot be divorced either from findings in other areas or from a sensitivity to the need for interdisciplinary integration? In learning to think critically, one learns to transfer what one

has learned about the logic of questions in one field to logically similar questions in other fields. Typically this begins with a recognition of the need to ask questions based on logical parallels between all fields of study, for example, skilled practice in questioning concepts and theories, in questioning data, in questioning the source or interpretation of data, in questioning the nature or organization of data, in questioning inferences, in questioning assumptions, in questioning implications and consequences, in questioning points of view and frames of reference, etc.

CRITERION # 7

Can it be used to assess those versatile and fundamental skills essential to being a responsible, decision-making member of the work-place? Critical thinking skills and abilities are highly transferable to the work-place. Since in learning to think critically we learn to take increasing charge of our minds as an instrument of learning — for example, reading, writing, speaking, and listening with greater discipline and skill — we are well situated to engage in collective problem solving and goal attainment, wherever they occur. The kind of "work" increasingly required in industry and business is "intellectual", that is, it requires workers to define goals and purposes clearly, seek out and organize relevant data, conceptualize those data, consider alternative perspectives, adjust thinking to context, question assumptions, modify thinking in the light of the continual flood of new information, and reason to legitimate conclusions. Furthermore, the intellectual work required must increasingly be coordinated with, and must profit from the critique of, fellow workers. There is no avoiding the need, therefore, to express ideas well, accurately represent and consider fairly the ideas of others, write clear and precise memos and documents, and coordinate and sequence all of these so that well-reasoned policies and decisions can be accurately understood and effectively implemented.

CRITERION # 8

Can it generate clear concepts and well-thought-out, rationally articulated goals, criteria, and standards? Since critical thinking is based on the art of monitoring one's thinking with standards implicit in the universal structure of thought, and since the use of these standards is implicit in intellectual history from Socrates through Einstein, there is no problem using critical thinking to generate clear concepts for testing, as well as rationally articulated goals, criteria, and standards.

CRITERION # 9

Can it account for the integration of adult-level communication skills, problem-solving, and critical thinking, and legitimately assess all of them without compromising essential features of any of them? Shallow concepts of critical thinking often distinguish critical thinking from problem solving and decision making as well as from reading, writing, and speaking skills. Once one considers a rich, substantive concept of critical thinking, however, it is clear that each of the basic skills of critical thinking are presupposed by each of the other skills, just as each of them is deeply interrelated to critical thinking as a whole. Consider, does it make sense to analyze potential solutions to problems or the implications of choosing an alternative in making a decision without using critical thinking? Clearly not. Every problem to be solved (or question to be settled) requires a *critical* analysis of the conditions under which it can be solved or settled. We, as problem-solvers, need to look critically at the purpose for solving the problem, we need to critically examine contextual factors, our assumptions, our concepts, what we are using as data, our organization of the data, the source of the data, our reasoning, the implications of our reasoning, our point of view, objections from other points of view. All of these are essential to higher order problem solving and decision making. Furthermore, all of these intellectual abilities are crucial to higher order reading, writing, speaking, and listening. To read must we analyze the text and re-create its logic in our own minds. To write we must construct a logic our readers can translate into the logic of their thought. To speak we must articulate our thoughts in such a way that our audience can translate our thoughts into their experiences. To listen we must analyze the logic of the thinking of the speaker. Intellectually disciplined reading, writing, speaking, and listening require, in other words, that we work explicitly with the logic we are constructing or re-constructing, using our grasp of the standards of critical thinking to communicate accurately and precisely, effectively solve problems, and rationally make decisions.

CRITERION # 10

Does it respect cultural diversity by focusing on the common-core skills, abilities and traits useful in all cultures? As the criterion presupposes, we can respect cultural diversity best by constructing tests in higher order thinking that focus on skills and abilities necessary in all modern cultures. In this way we can legitimately justify assessing it in all cultural groups. Basic critical thinking skills and abilities — because they are based on fundamental elements implicit in the structure of all reasoned thought *per se*, and because their mastery is essential to higher order thinking in all academic, professional, personal, and public life — are an appropriate foundation for assessment.

CRITERION # 11

Does it test for thinking that promotes (to quote the September, 1991 Kappan) "the active engagement of students in constructing their own knowledge and understanding?" Narrow concepts of critical thinking sometimes characterize it in negative terms, as a set of tools for detecting mistakes in thinking. A rich, substantive concept of critical thinking, however, highlights its central role in all rationally defensible thinking, whether that thinking is focused on assessing thought or products already produced, or actively engaged in the construc-

tion of new knowledge or understandings. Well-reasoned thinking, whatever its end, is a form of creation and construction. It devises and articulates purposes and goals, translates them into problems or questions, seeks data that bear upon problems or questions, interprets those data on the basis of concepts and assumptions, and reasons to conclusions within some point of view. All of these are necessary acts of the reasoning mind and must be done "critically" to be done well. Hence all require critical thinking.

CRITERION # 12

Does it concentrate on assessing the fundamental cognitive structures of communication? Each of the dimensions identified in the objective is either straightforwardly a critical thinking ability or depends on a critical thinking ability. The writer's or speaker's *purpose, implications, assumptions, point of view,* etc., are all elements of thought, and the ability to identify and assess those as one reads or listens — the ability to construct in one's mind an accurate and fertile interpretation — is simply thinking *by* listening, thinking *by* reading.

A similar reliance on elements of thought is central to writing or speaking effectively at any educational level. The knowledge of how to gather and present evidence, to make clear one's own assumptions, to see the implications of a position: these are critical thinking abilities.

All forms of communication, moreover, rely on critical thinking standards. Essays and interpretations of essays, utterances and interpretations of utterances, need to be *relevant, logical, consistently* worked out; evidence needs to be recorded and reported *accurately;* points need to be made *clearly* and with as much *precision* as the subject permits; topics need to be covered in *depth* and presented *fairly.*

CRITERION # 13

Can it be used to assess the central features of making rational decisions as a citizen, a consumer, and a part of a world economy? Both public and private life increasingly require mastery of the basic skills and abilities of critical thinking. When this mastery is absent the public degenerates into a mass society susceptible to manipulation by public relations specialists who can engineer political victories by an adroit use of mud slinging, scare tactics, shallow nationalism, fear, envy, stereotypes, greed, false idealism, and maudlin sentimentality. Modern citizenship requires basic critical thinking skills and abilities throughout. The modern citizen should be able to assess the arguments presented for his or her assent, must rationally adjudicate between conflicting points of view, must attempt to understand a culturally complex world, must assess the credibility of diverse sources of information, must translate between conflicting points of view and diverse appeals, must rationally decide priorities, must seek to understand complex issues that involve multiple domains (for example, the environmental, moral, economic, political, scientific, social, and historical domains). Without a solid grounding in critical thinking, citizens are intellectually disarmed, incapable of discharging their civic responsibilities or rationally exercising their rights.

CRITERION # 14

Can it avoid reducing a complex whole to oversimplified parts? Testing for a rich, substantive concept of critical thinking is testing for skills of reasoning in terms of elements of thought, for the ability to orchestrate those elementary skills, for the affective dimensions that make critical thinking actualizable in practice, and for universal intellectual standards, in short for a rich and complex whole rather than for fragmented parts.

CRITERION #15

Can it articulate what is central to basic skills for the future? Basic skills are constituted by the structures explicated in a rich, substantive concept of critical thinking. To teach reading is to teach the ability not merely to repeat content, but to reconceptualize that content, to see applications of the main ideas, to generalize from them, critique them, see them in context, to enter with empathy into another's point of view. To teach writing as a basic skill is to teach not merely grammar and punctuation, but the ability to arrange one's ideas logically and consistently, to anticipate reasonable objections, to transfer ideas to the page in a way that makes them decipherable in all their complexity by a reader. To teach math as a basic skill is not primarily to teach how to solve pre-selected, individual, isolated problems out of context, but to teach the ability to begin to make sense of the world mathematically, to think quantitatively, to be able to see mathematical patterns, to set up the construction of problems and then creatively go about solving them. Critical thinking abilities like these do not exist somehow *in addition to* the basic skills of life; they *constitute* the basic skills of life.

CRITERION #16

Can it provide the kind of skills that are seen as valuable outside the school as well as inside it? Critical thinking provides skills that are seen as valuable by practitioners of the academic disciplines, by responsible leaders of government, of the professions, of business, by citizens interested in their environmental, physical, and economic welfare. In all such areas what is needed are ways to adapt to rapidly changing knowledge, to recognize problems and see their implications before they become acute, to formulate approaches to their solution that recognize legitimately different points of view, to draw reasonable conclusions about what to do. Increasingly, one is hearing statements such as the one made by David Kennedy, the president of Stanford University, to 3,000 college and university presidents:

> It simply will not do for our schools to produce a small elite to power our scientific establishment and a larger cadre of workers with basic skills to do routine work. Millions of people around the world now have these same basic skills and are willing to work twice as long for as little as one-tenth our basic wages. To maintain and enhance our quality of life, we must develop a leading-edge economy based on

workers who can think for a living. If skills are equal, in the long run wages will be too. This means we have to educate a vast mass of people capable of thinking critically, creatively, and imaginatively.

CRITERIA #17 AND #18

Can critical thinking be assessed in a way that requires evaluation of authentic problems in realistic contexts, where the abilities assessed include those of formulating the problem and initial screening of plausible solutions? Yes. Testing of authentic skills, abilities and dispositions in authentic contexts can be accomplished by using a combination of *a)* standard multiple-choice items, *b)* machine-gradable multiple-rating items and *c)* short essay items.

a) The standard multiple-choice part of the assessment would be an expanded version of established critical thinking tests, such as the Watson-Glaser or Cornell tests. It is suitable for assessing micro-dimensional critical thinking skills, like identifying the most plausible assumption, recognizing an author's purpose, selecting the most defensible inferences, and such like.

b) The multiple-rating part of the assessment would test more open-ended and larger-domained abilities, like thinking within opposing points of view, the willingness to suspend judgment, the ability to synthesize disparate data into a logical scheme, to take established findings and generalize them into new contexts, etc.

The multiple-rating portion of the assessment, to be reliable, must:

i) embody a rich and substantive idea of critical thinking,

ii) be constructed and monitored by critical thinking experts who have such a concept,

iii) be changed often (5% annually) to assess critical thinking with respect to authentic contemporary issues.

c) The essay part of the assessment would be designed to address critical thinking abilities and traits that involve creating a logic to capture a situation rather than selecting from among possibilities suggested by the test. Examples include the ability to construct an interpretation, to make a logical outline of a text, to figure out ways to gather information, to take an unclear and complex real issue and reformulate it so as to make it more amenable to solution.

Validity on the essay part of the assessment requires that the test be:

i) constructed by experts in critical thinking,

ii) assembled from a large and rotating bank of short essay questions to allow for items that show no significant differences,

iii) centrally graded by teams well-trained in a full concept of critical thinking in order to assure quality control.

CRITERION #19

Can critical thinking be assessed nationally in a way that is financially affordable? To make it affordable, the constructed response segment of the assessment should be administered not to the population of students as a whole, but rather to a representative sample of the student population of a school system. The assessment should be (a) paid for by school systems that contract to have their students tested, and (b) constructed, monitored, administered, and graded by a private agency with critical thinking credentials, or at least under the direction of scholars with a solid grounding in the research into critical thinking.

CRITERIA #20 AND #21

Can critical thinking be assessed so as to gauge the improvement of students over the course of their education and to measure the achievement of students against national standards? To evaluate students in both these dimensions requires:

a) an assessment administered as a pre-test at the 6th grade and then as a follow up at the 9th and 12th grades (to provide for value-added judgments).

b) a criterion-referenced assessment that is built on clear, consistently applied quality-norms that are derived from a rich and substantive concept of critical thinking (to provide for the measuring of national progress).

✦ *What, Specifically, Are the Dangers of a Non-Substantive Concept of Critical Thinking?*

It is important to be alert to the dangers posed by a non-substantive concept of critical thinking. Such a concept exists when, separate from a consideration of the research in the field, a person or institution presupposes *a)* that the meaning or terminology of critical thinking is intuitively obvious (hence not in need of scholarly analysis), or *b)* that each concept underlying critical thinking (such as assumption, inference, implication, reasoning, ...) can be analyzed separately from a theory that accounts for the interrelation of these concepts, or *c)* that the skills of critical thinking can be adequately cultivated without reference to the values, traits of mind, and dispositions that underlie those skills.

1) There are at least three serious problems that may result from the use of a theoretically superficial concept of critical thinking:

a) important critical thinking concepts, which must be clearly defined to be used effectively in assessment, may be used vaguely, inconsistently, incorrectly, or misleadingly,

b) a false, misleading, or simplistic over-arching concept of critical thinking may be fostered, or

c) an unrealistic strategy for the assessment and cultivation of critical thinking may be incorporated into testing and teaching.

Many examples of the unwitting use of a non-substantive concept of critical thinking could be cited — such as "thinking skills" programs devoid of intellectual standards (which, for example, systematically confuse "inferences" with "valid inferences" and "analogies" with "sound analogies"), or testing personnel who lack adequate grounding in critical thinking theory (and so, for example, frequently confuse assumptions with inferences or inferences with implications). The most far-reaching danger occurs when influential educational systems or institutions, like state departments of education, inadvertently incorporate a non-substantive concept of critical thinking into statewide curriculum standards or into statewide testing programs. This can result in significant, unintended negative consequences, for example: thousands of teachers encouraged to follow a misconceived model for the assessment of reasoning, leading to mis-instruction on a grand scale.

2) Illustration: The California Direct Writing Assessment

We shall look at one important case. Unfortunately, given the brevity of this paper, one case must stand for all. The case we have chosen concerns the *Integrated Language Arts Assessment of the California Assessment Program*, a massive statewide program that has impact not only on every student in the public schools of California, but also, because of the leadership role of California in assessment, on national teaching and testing practices as well. It appears that three fundamental mistakes occurred in the design of the direct writing assessment:

a) Though one of the goals of the program was to place an emphasis on the quality of reasoning and critical thinking in writing, it appears that no one with a research background in critical thinking reviewed the articulation or implementation of the assessment prompts. (We infer this from the fact that fundamental conceptual errors occur both in the prompts themselves and in the application of criteria to student constructed responses.)

b) It was assumed, inappropriately, that classroom teachers without extended training in critical thinking are able to effectively assess student essays that call for evaluative reasoning. We infer this from statements descriptive of the assessment design like:

> Teachers on the CAP writing Development Team develop all the testing and instructional materials for assessment. For every type of writing assessed, the team develops a special set of prompts ... and a scoring guide that identifies the thinking and writing requirements for that type of writing....
>
> Essays are scored in four to six days by several hundred teachers at four regional scoring centers. A special handbook for each grade level provides teachers with practical instructional materials for each type of writing, including sample prompts, illustrative essays, and related readings.

c) The resulting assessment was not monitored by anyone with a research background in critical thinking. (We infer this from the fact that model "strong" answers purporting to illustrate critical reasoning are showcased that are in fact patently very weak answers, containing virtually no reasoning at all.)

Consider Figure 1 and Figure 2 used as illustrations of the nature and quality of the writing assessment program in an article authorized and developed by the staff of the California Assessment Program. It is entitled "California: The State of Assessment" and was written for an important national anthology, *Developing Minds* (more than 150,000 copies disseminated by ASCD). The show-piece article, in which these figures occur, argues that the examples illustrate a "state-of-the-art teacher-developed writing assessment" that is sophisticated in "its testing, scoring, and reporting systems" and designed to "include only those tasks that will stimulate high-quality instruction."

There are a number of problems illustrated in these figures that a substantive understanding of critical thinking would have avoided:

1) A description of subjective reactions was systematically confused with sound evaluative reasoning. It is important to distinguish questions like, "Is rock music good music?" or "Does rock music excel as a form of music?" (which call for objective evaluation) from questions like, "Do you enjoy rock music?" or "Does rock music stir powerful emotions in you?" (which call, not for reasoning, but for the description of subjective reactions). Apparently the test developers were unclear about this distinction.

2) The assessing teachers did not notice that the student failed to respond to the directions. The student did not develop evaluative reasoning, did not support his judgment with reasons and evidence, did not consider possible criteria on which to base his judgment, did not analyze the subject in the light of the criteria, and did not select evidence that clearly supported his judgment. Instead the student described an emotional exchange, asserted — without evidence — some questionable claims, and expressed a variety of subjective preferences (a fuller critique of the student essay is available in Chapter 8, "Why Students — and Teachers — Don't Reason Well"). The assessing teachers were apparently too confused about the nature of evaluative reasoning or the basic notions of criteria, evidence, reasons, and well-supported judgment to notice the discrepancy.

3) The California State Department of Education assessment staff did not notice these errors once they were made. Instead of catching the errors once made, the California Department of Education chose to use the mis-graded student essay *as a showcase model* to disseminate

BEST COPY AVAILABLE

Evaluative Essay Sample

EVALUATION. Students were asked to write an evaluative essay, make judgments about the worth of a book, television program, or type of music and then support their judgments with reasons and evidence. Students must consider possible criteria on which to base an evaluation, analyze their subject in light of the criteria, and select evidence that clearly supports their judgments. Each student was assigned one of the following evaluative tasks:

- To write a letter to a favorite author telling why they especially liked one of the author's books.
- To explain why they enjoyed one television program more than any others.
- To justify their preference for a particular type of music.

The tasks made clear that students must argue convincingly for their preferences and not just offer unsupported opinions.

This is a sample essay from a student who demonstrated exceptional achievement.

Rock Around the Clock

"Well, you're getting to the age when you have to learn to be responsible!" my mother yelled out.

"Yes, but I can't be available all the time to do my appointed chores! I'm only thirteen! I want to be with my friends, to have fun! I don't think that it is fair for me to baby-sit while you go run your little errands!" I snapped back. I sprinted upstairs to my room before my mother could start another sentence. I turned on my radio and "Shout" was playing. I noted how true the song was and I threw some punches at my pillow. The song ended and "Control" by Janet Jackson came on. I stopped beating my pillow. I suddenly felt at peace with myself. The song had slowed me down. I pondered briefly over all the songs that had helped me to control my feelings. The list was endless. So is my devotion to rock music and pop rock. These songs help me to express my feelings, they make me wind down, and above all they make me feel good. Without this music, I might have turned out to be a violent and grumpy person.

Some of my favorite songs are by Howard Jones, Pet Shop Boys, and Madonna. I especially like songs that have a message in them, such as "Stand by Me", by Ben E. King. This song tells me to stand by the people I love and to not question them in times of need. Basically this song is telling me to believe in my friends, because they are my friends.

My favorite type of music is rock and pop rock. Without them, there is no way that I could survive mentally. They are with me in times of trouble, and best of all, they are only a step away.

California classroom teachers wrote comments like these after reading and scoring students' evaluative essays:

- "Evidence of clear thinking was heavily rewarded in our scoring."
- "I am struck by how much some students can accomplish in 45 minutes; how well they can sometimes marshal the ideas; and with how much flair and sparkle they can express themselves."
- "More emphasis should be placed on critical thinking skills, supporting judgments, and tying thoughts and ideas together. Far too many papers digress, summarize, underdevelop, or state totally irrelevant facts."
- "Students generally need to develop skills in giving evidence to support their judgments. I plan to spend more time on these thinking skills next year."

Source: California State Department of Education, 1988.

Figures 1 and 2 come from "California: The State of Assessment", Anderson, Robert L. in *Developing Minds*, edited by Art Costa, pp. 314–25.

figure 1

nationally as illustrating "exceptional achievement" in reasoned evaluation, and as a *model* of their assessment of reasoned writing. We conclude that the California Assessment Program does not use scholars with a background in critical thinking research, any of whom would surely have recognized the problem.

Fundamental misconceptions of the nature of critical thinking and reasoned discourse, such as those documented above, must not be replicated in a national assessment program. Steps should be taken to insure that a substantive concept of critical thinking and a well-supervised implementation of that concept form the basis of the finished assessment program.

CAP Grade 8 Direct Writing Assessment
Achievement in Evaluation

Score Point	Percentage of California Grade 8 Students*	Cumulative Percentage	Description of Achievement
6 Exceptional Achievement	0.5		The student produces convincingly argued evaluation; identifies a subject, describes it appropriately, and asserts a judgment of it; gives reasons and specific evidence to support the argument; engages the reader immediately, moves along logically and coherently, and provides closure; reflects awareness of reader's questions or alternative evaluations.
5 Commendable Achievement	8.1	8.6	The student produces well-argued evaluation; identifies, describes, and judges its subject: gives reasons and evidence to support the argument; is engaging, logical, attentive to reader's concern; is more conventional or predictable than the writer of a 6.
4 Adequate Achievement	25.5	34.1	The student produces adequately argued evaluation; identifies and judges its subject; gives at least one moderately developed reason to support the argument; lacks the authority and polish of the writer of a 5 or 6; produces writing that, although focused and coherent, may be uneven; usually describes the subject more than necessary and argues a judgment less than necessary.
3 Some Evidence of Achievement	42.4	76.5	The student states a judgment and gives one or more reasons to support it; either lists reasons without providing evidence or fails to argue even one reason logically or coherently.
2 Limited Evidence of Achievement	19.2	95.7	The student states a judgment but may describe the subject without evaluating it or may list irrelevant reasons or develop a reason in a rambling, illogical way.
1 Minimal Evidence of Achievement	3.6	99.3	The student usually states a judgment but may describe the subject without stating a judgment; either gives no reasons or lists only one or two reasons without providing evidence; usually relies on weak and general personal evaluation.
No response		0.3	
Off Topic		0.5	

*This column does not total to 100% because of rounding

figure 2

✦ *Section Three: The Four Domains of Critical Thinking*

What are the four component domains of critical thinking and their implications for the assessment of higher order thinking?

✦ *Elements of Thought*

As soon as we move from thought which is purely associational and undisciplined, to thinking which is conceptual and inferential, thinking which attempts in some intelligible way to figure something out, to use the power of reason, then it is helpful to think about what can be called "the elements of thought." The elements of thought are the basic building blocks of thinking, essential dimensions of reasoning whenever and wherever it occurs. Working together, they shape reasoning and provide a general logic to reason. We can articulate these elements by paying close attention to what is implicit in the attempt on the part of the mind to figure anything out whatsoever. Once we make them clear, it will be obvious that each of them can serve as an important touchstone or point of assessment in critical analysis and in the assessment of thinking.

For each of the elements of thought there is a cluster of attendant basic thinking skills. Because they involve fundamental structures of thought, these skills can be characterized as micro-skills, those skills out of which larger-domained critical thinking abilities are built. Being able to think critically about a particular issue, then, will include the ability to identify, clarify, and argue for and against alternative formulations of the elements of thought.

The basic conditions implicit whenever we gather, conceptualize, apply, analyze, synthesize, or evaluate information — the elements of thought — are as follows:

1) *Purpose, Goal, or End in View.* Whenever we reason, we reason to some end, to achieve some objective, to satisfy some desire or fulfill some need. One source of problems in reasoning is traceable to defects at the level of goal, purpose, or end. If the goal is unrealistic, for example, or contradictory to other goals we have, confused or muddled in some way, then the reasoning used to achieve it is problematic.

   An assessment of critical thinking, then, would test, at the appropriate educational level, skills of being able to state an author's purpose, to identify a plausible statement of an author's goals from a list provided, to rank formulations of an author's objectives

according to which are more or less reasonable in light of a particular passage, to distinguish clearly between purposes, consequences, assumptions, and other elements of thought.

2) *Question at Issue, or Problem to be Solved.* Whenever we attempt to reason something out, there is at least one question at issue, at least one problem to be solved. One area of concern for reasoners, therefore, will be the formulation of the question to be answered or problem to be solved, whether with respect to their own reasoning or to that of others.

   Assessing skills of mastery of this element of thought would test students' ability to formulate a problem in a clear and relevant way, to choose from among alternative formulations, to discuss the merits of different versions of the question at issue, to recognize key common elements in statements of different problems, to structure the articulation of problems so as to make possible lines of solution more apparent.

3) *Point of View, or Frame of Reference.* Whenever we reason, we must reason within some point of view or frame of reference. Any "defect" in that point of view or frame of reference is a possible source of problems in the reasoning. A point of view may be too narrow, too parochial, may be based on false or misleading analogies or metaphors, may contain contradictions, and so forth.

   Levels of skill here would be tested with reference to being able to enunciate an author's point of view in a passage, to adjudicate between different statements of that point of view, to recognize bias, narrowness, and contradictions when they occur in the point of view, to recognize relations between the frame of reference being used and its implications, assumptions, and main concepts.

4) *The Empirical Dimension of Reasoning.* Whenever we reason, there is some "stuff", some phenomena about which we are reasoning. Any "defect", then, in the experiences, data, evidence, or raw material upon which a person's reasoning is based is a possible source of problems.

   Students would be tested, again, based on their level, on their ability to distinguish evidence from conclusions based on that evidence, to give evidence themselves, to identify from a pre-selected list data that would support an author's positions, data that would oppose it, data that would be neutral, to notice the presence or lack of relevant evidence, to recognize, to be intellectually courageous in recognizing (and labeling as such) mere speculation that goes beyond the evidence.

5) *The Conceptual Dimension of Reasoning.* All reasoning uses some ideas or concepts and not others. These concepts can include the theories, principles, axioms and rules implicit in our reasoning. Any "defect" in the concepts or ideas of the reasoning is a possible source of problems.

   The assessment of the relevant higher order thinking would test the ability to identify main concepts of a passage, to choose among different versions of those concepts (some perhaps equally good), to see relations among concepts, to reason about the similarity of points of view on the basis of similarity of fundamental concepts, to distinguish central from peripheral concepts, derived concepts from basic concepts, to see the implications of using one concept rather than another.

6) *Assumptions.* All reasoning must begin somewhere, must take some things for granted. Any "defect" in the assumptions or presuppositions with which the reasoning begins is a possible source of problems.

   Assessing skills of reasoning about assumptions would test the ability to identify assumptions underlying given inferences, points of view, and goals, to evaluate the accuracy of different formulations of the assumptions, to distinguish between assumptions and inferences, to rank assumptions with respect to their plausibility, to be intellectually fairminded by choosing the most plausible version of assumptions underlying points of view with which they disagree.

7) *Implications and Consequences.* No matter where we stop our reasoning, it will always have further implications and consequences. As reasoning develops, statements will logically be entailed by it. Any "defect" in the implications or consequences of our reasoning is a possible source of problems.

   Skills to be assessed would include the ability to identify important implications, to do so by selecting from a list of possible implications, to make fine discriminations among necessary, probable, and improbable consequences, to distinguish between implications and assumptions, to recognize the weakness of an author's position as shown by the implausibility of its implications, to exercise intellectual fairmindedness in discriminating between the likelihood of dire and mild consequences of an action to which one is opposed.

8) *Inferences.* Reasoning proceeds by steps in which we reason as follows: "Because this is so, that also is so (or probably so)," or "Since *this*, therefore *that*." Any "defect" in such inferences is a possible problem in our reasoning.

Assessment would test, in a way geared to their educational level, students' ability to recognize faulty and justified inferences in a passage, to rank inferences with respect to both their plausibility and their relevance, to make good inferences in their own reasoning, to discriminate among various formulations of an author's inferences with respect to which is most accurate, to take something they do not believe but to entertain it for the sake of argument and draw reasonable inferences from it.

Assessment of Elements of Thought

Any program for the assessment of critical thinking skills must itself be assessed in terms of its validity and reliability in testing for the ability to think about, and in terms of, the elements of thought. These abilities can be successfully assessed in three related ways: by a restricted use of standard multiple-choice items, by multiple-rating items, and by short essay items. Both multiple-choice and multiple-rating items are machine-gradable, while essay items are not.

Although our recommendations about the content of the assessment will be spelled out in detail in Section Four, some of these can be anticipated here with respect to the assessment of reasoning abilities centering around the elements of thought.

BEST COPY AVAILABLE

Multiple-choice testing (as in the existing *Watson-Glaser Critical Thinking Appraisal* or the *Cornell Critical Thinking Tests*) is an important part of an assessment of critical thinking, but its legitimate use is restricted to testing only the most basic skills of identifying and recognizing elements of thought, and then only as they occur in relatively short and unambiguous excerpts.

Within this domain, multiple-choice questions will require students:

- to identify an author's purpose in a passage;
- to rate selected inferences as justified, probably true, insufficiently evidenced, probably false, unjustified;
- to select among formulations of the problem at issue in a passage those that are clearly reasonable, probably reasonable, probably unreasonable, clearly unreasonable;
- to recognize unstated assumptions;
- to distinguish evidence from hypotheses and conclusions;
- to rate described evidence as reliable, probably reliable, probably not reliable, unreliable.

✦ *Abilities*

The elements of thought do not exist in isolation from one another, nor — more importantly for the concept of an assessment procedure — do they exist outside a particular context of application. In the practice of good critical thinking, skills more closely associated with elements of thought are orchestrated into larger-domained abilities which are applied to thinking about complex and sometimes ambiguous issues, problems, decisions, theories, states of affairs, social institutions, and human artifacts.

These critical thinking abilities include being skillful at:

1) refining generalizations and avoiding over-simplifications,
2) comparing analogous situations: transferring insights into new contexts,
3) developing one's perspective: creating or exploring the implications of beliefs, arguments, or theories,
4) clarifying issues, conclusions, or beliefs,
5) clarifying and analyzing the meanings of words and phrases,
6) developing criteria for evaluation: clarifying values and standards,
7) evaluating the credibility of sources of information,
8) questioning deeply: raising and pursuing root or significant questions,
9) analyzing or evaluating arguments, interpretations, beliefs, or theories,
10) generating or assessing solutions,
11) analyzing or evaluating actions or policies,
12) reasoning dialogically: comparing perspectives, interpretations, or theories,
13) reasoning dialectically: evaluating perspectives, interpretations, or theories,
14) reading critically: constructing an accurate interpretation of, understanding the elements of thought in, and evaluating, the reasoning of a text,
15) listening critically: constructing an accurate interpretation of, understanding the elements of thought in, and evaluating, the reasoning of an oral communication,
16) writing critically: creating, developing, clarifying, and conveying, in written form, the logic of one's thinking,
17) speaking critically: creating, developing, clarifying, and conveying, in spoken form, the logic of one's thinking.

Abilities like these play a central role in a rich and substantive concept of critical thinking. They are essential to approaching actual issues, problems, and situations rationally. Understanding the rights and duties of citizenship, for example, requires that one at least have the ability to compare perspectives and interpretations, to read and listen critically, to analyze and evaluate policies. In fact, there is no macro-ability on the list that would not be relevant or even crucial to thinking deeply about the rights and duties of citizenship. Similarly, the capacity to make sound decisions, to participate knowledgeably in the work-place, to function as part of a global economy, to master the content in anything as complex as the academic disciplines, to apply those subject area insights to real-life situations, to make insightful cross-disciplinary connections, to communicate effectively — each of these relies in a fundamental way on having a significant number of the abilities listed. Take, for example, the capacity to make sound decisions: such decision-making is hardly possible without an attendant ability to (going down the list of abilities in order) refine generalizations, compare analogous situations, develop one's perspective, clarify issues, and so forth.

The last four abilities listed — the ability to read, write, listen, and speak, each in a critical, informed, constructive way — are best considered not as in the usual model, not as manifestations of thinking already accomplished, but as being themselves actual modes of constructive thinking. As such, they are structured amalgams of elementary skills together with any number of other abilities.

BEST COPY AVAILABLE

ASSESSMENT OF ABILITIES

The assessment of abilities, too often neglected, is essential to assessment of critical thinking. Since these *are* the abilities implicit in the realistic use of thinking, no assessment tool that fails to assess a significant number of these abilities could justifiably be called an assessment of higher order thinking. The assessment, moreover, needs to address such abilities *directly* (rather than through secondary indicators), *systematically* (rather than haphazardly as a result of an attempt to assess other variables like academic achievement), and in settings as *authentic* as possible given the requirement of *uniform, relevant grading*.

Assessment of abilities that meets these four criteria cannot be accomplished within the confines of a standard multiple-choice-type test. It can be accomplished, however, for all of the abilities (except those having to do with oral communication), by means of a combination of machine-gradable multiple-rating items and essay items.

For any macro-ability, there will be dimensions of the ability that are *generative* and other dimensions of it that are selective. In trying to solve a real problem, for example, much of one's thinking is devoted to *generating* a formulation of the problem that will make it more susceptible to solution. Another, and quite different, aspect of problem solving, is the ability to *select*, from among a large variety of possibilities, that avenue of thought which will most likely result in a solution. Students who are trained using a rich, substantive concept of critical thinking tend to improve in both dimensions of this ability, and both are genuine dimensions of real problem-solving.

The selective dimensions of an ability can be assessed accurately, even in complex, ambiguous, and subtle cases, using multiple-rating items. The generative dimension, on the other hand, cannot. Since it requires students to come up with their own critical thinking approaches within that macro-ability, this dimension can be assessed adequately only by carefully constructed and carefully graded essay tests. Details of the assessment and samples of assessment items will be presented in Section Four.

✦ *Affective Dimensions*

Higher order thinking requires more than higher order thinking *skills*. Critical thinking, in any substantive sense, includes more than abilities. The concept also includes, in a crucial way, certain attitudes, dispositions, passions, traits of mind. These affective dimensions are not merely important to critical thinking, they are essential to the effective use of higher order thinking in real settings.

These affective dimensions include:

1) thinking independently,
2) exercising fairmindedness,
3) developing insight into egocentricity and sociocentricity,
4) developing intellectual humility and suspending judgment,
5) developing intellectual courage,
6) developing intellectual good faith and integrity,
7) developing intellectual perseverance,
8) developing confidence in reason,
9) exploring thoughts underlying feelings and feelings underlying thoughts,
10) developing intellectual curiosity.

Without *intellectual perseverance*, one could not solve the complicated, multi-faceted problems one confronts in industry. Without *intellectual courage*, one could not maintain a defense of citizenship rights in the face of scare tactics. Without *fairmindedness*, one could not enter into another's point of view and thus would lack that empathetic understanding necessary for a reasonable approach to living in a pluralistic society. Without *developing insight into egocentricity and sociocentricity* one could employ one's reasoning skills in a merely self-serving and prejudiced way. Without *confidence in reason* one could not adequately address those complex and frequently ambiguous real-life problems that require reasonable decisions in the face of crucial uncertainties.

ASSESSMENT OF AFFECTIVE DIMENSIONS

The assessment of affective dimensions of critical thinking is an important part of an assessment of higher order thinking. An initial problem is that from the fact that all these dimensions are essential, it does not follow that all are directly testable, nor does it follow that *any* of them is *easily* testable. For some of these affective dimensions (intellectual perseverance, for example), any testing would have to take place over an appropriately long period of time and thus could not be legitimately assessed at all during a time-frame suitable for a national test.

Nevertheless, a number of affective dimensions can be assessed in a relatively straightforward way using essay items and, especially, machine-gradable multiple-rating items.

"Reasoning Within Conflicting Points of View," a central aspect of the disposition of fairmindedness, is already being assessed on the revised version of the *Watson-Glaser Critical Thinking Appraisal*. This section of the *Appraisal* asks students to select the strongest (that is, the most defensible) argument in favor of each side of a pair of conflicting and sometimes emotionally charged points of view. Proficiency on these items indicates a fairminded willingness to distinguish the concept of *reasonable defensibility* from that of *personal belief*.

Multiple-rating items are currently being prepared that address aspects of intellectual courage, other aspects of fairmindedness, aspects of intellectual humility, and aspects of the development of insight into one's own egocentricity and sociocentricity.

✦ *Intellectual Standards*

In any domain where assessment is taking place, there are standards implicit in that assessment. Higher order thinking is thinking that meets universal intellectual standards. Thus, when assessing a student's ability to compare and evaluate perspectives (a macro-ability) and to do so with fairmindedness (a trait of mind), we would judge whether she had made such evaluations in a *relevant* and *consistent* way, with attention to *accuracy*, *fairness*, and *completeness* in describing each perspective, and with a sensitivity to the degree of *precision* appropriate to the topic. We would assess critical thinking about and in terms of the elements of thought in very much the same way: to judge a person's skill at recognizing the frame of reference underlying a position, we would want to judge whether she could see *relevant* alternatives, whether the frame of reference she identified fits the available *evidence*, whether her answer was *deep* or merely mechanical, *clear* or vague, *fair* or biased. Intellectual standards apply to thinking in every subject.

The process of learning to teach so as to foster critical thinking is the very process by means of which one establishes intellectual standards for assessing thinking, and, by extension, for assessing instruction itself.

Such standards are more useful if they are made explicit — to the students who are taking the test, to those doing the assessing, and to classroom teachers. Making standards explicit benefits student test-takers because they can then see that there are standards, that the standards are not arbitrary, and that understanding the standards gives them insight into what good critical thinking is. It benefits those doing the assessing because, in addition to the reasons already mentioned, it fosters both a uniformity in grading and a strong correlation between the grade and the skills being graded. Judging a response by how *clearly* and *completely* it states a position, for example, is using a critical thinking standard and dictates a certain level of assessment; judging a response by how *concisely* or how *elegantly* it states a position, on the other hand, is using a standard that is inappropriate to critical thinking assessment. Explicit standards — part of a rich and substantive concept of critical thinking — might have avoided at least some of the mistaken assessment on the *California Assessment Program*, cited earlier (see p. 94). Thus, making standards explicit promotes both the reliability and the validity of the assessment. Finally, it benefits classroom teachers because such standards can readily be built into classroom instruction. The standards, after all, are those implicit in teaching for higher order thinking; they are therefore invaluable both for teachers to use explicitly with their classes and — an essential feature of critical thinking-internalized — for students to learn to use as part of assessing themselves.

Intellectual Standards
That Apply to Thinking in Every Subject

Thinking that is:		*Thinking that is:*
Clear	vs	Unclear
Precise	vs	Imprecise
Specific	vs	Vague
Accurate	vs	Inaccurate
Relevant	vs	Irrelevant
Plausible	vs	Implausible
Consistent	vs	Inconsistent
Logical	vs	Illogical
Deep	vs	Superficial
Broad	vs	Narrow
Complete	vs	Incomplete
Significant	vs	Trivial
Adequate *(for purpose)*	vs	Inadequate
Fair	vs	Biased or One-Sided

✦ Section Four: Recommendations of the Center for Critical Thinking

What is the most workable solution to the design of a process to assess higher order thinking?

In this section we will *1)* briefly survey existing assessment tools; *2)* make recommendations regarding the substance and format of a national assessment tool — the critical thinking domains to be assessed, the varieties of assessment strategies to be used (including sample test items), and the dual *inter*disciplinary and *intra*disciplinary scope of the assessment — *3)* appraise the value of the proposed assessment strategy for the reform of instruction, and *4)* make recommendations regarding the implementation of the assessment.

✦ Existing Assessment Tools

There are limitations in all twelve of the commercially available critical thinking tests as instruments for assessing higher order thinking:

Cornell Class Reasoning Test, Form X (1964)
Cornell Conditional Reasoning Test, Form X (1964)
Cornell Critical Thinking Test, Level X (1985)
Cornell Critical Thinking Test, Level Z (1985)
The Ennis–Weir Critical Thinking Essay Test (1985)
Judgement: Deductive Logic and Assumption Recognition (1971)
Logical Reasoning (1955)
New Jersey Test of Reasoning Skills (1983)
Ross Test of Higher Cognitive Processes (1976)
Test on Appraising Observations (1983)
Test of Enquiry Skills (1979)
Watson–Glaser Critical Thinking Appraisal (1980)

In addition there are limitations in all of the other available "higher studies" tests which might be taken as a possible model for the assessing of higher order thinking: the SAT, LSAT, the Test of Academic Aptitude (British), ACT, the Graduate Record Exam, the Commonwealth Secondary Scholarships Exam (Australia). We do not have the space here to review each of these tests one-by-one. Instead we will summarize the general situation as we see it.

Though aspects and dimensions of critical thinking are tested, some more and some less, in all of the above tests, none has been designed with the 21 criteria in Sections one and two in mind. Most importantly, none was designed to serve as a national assessment tool which establishes national standards in higher order thinking and as a motivation for and guide to instruction.

Behind none of these tests was there a comprehensive model for the elements of thought, the abilities of critical thinking, or the affective dispositions (as we have here provided). The relative recentness of the bulk of scholarship in critical thinking makes it unlikely that long-established tests will fill the bill.

Of course any new test for assessing higher order thinking should be based on a thorough review of established test strategies to incorporate those with significant application.

Given the need for assessment on the basis of a rich and substantive concept of critical thinking, there are two areas where competing values and objectives come into play.

The first concerns the *substance and format* of the test itself: Which domains exactly are to be covered, and with what emphases? What kinds of question will be asked? Will it include both interdisciplinary and intradisciplinary items? What kind of assessment questions best test for skills of citizenship and the challenges of the work-place?

The second area concerns the *implementation* of the test and how it is conceived: Should it be value-added or simply criterion-referenced? Who will do the assessing and who will be assessed? How much will the assessment cost and who will pay for it? How often will the test be given?

Some of these are difficult questions, with genuine values and goals on different sides, where reasonable cases can be made for more than one position. Others of these questions are clearer, especially once the objectives of the test as a whole are brought into focus.

✦ Substance and Format

The overall recommendations of the Center for Critical Thinking are set forward below.

1) DOMAINS TO BE ASSESSED

The national assessment of higher order thinking must test for a rich and substantive concept of critical thinking, and this testing must be geared to assessment within all four domains of critical thinking.

a) Elements of Thought

Skills of identifying, explicating, and using the elements of thought need to be assessed. They are necessary for any of the abilities to be employed with precision, depth, or accuracy. They are required if essential affective traits are to be rooted in solid, locatable, intellectual skills and the concepts they presuppose.

Lack of a solid grounding in these skills, and the concepts behind them, results in thinking which, good intentions notwithstanding, is far removed from the close, careful reasoning demanded by the rigors of higher order thinking. Among testing personnel, lack of the informed use of these concepts is part of what results in such poor assessment tools and grading as we found in the *California Direct Writing Assessment*.

Critical thinking in students requires them to be able to perform well, with an expertise appropriate to their grade level, on items testing a list of skills that center around the elements of thought:

- identify a plausible statement of a writer's purpose;
- rank formulations of an author's objectives;
- distinguish clearly between purposes, consequences, assumptions, and inferences;
- choose the most reasonable statement of the problem an author is addressing;
- discuss reasonably the merits of different versions of the question at issue;
- recognize key common elements in formulations of different problems;
- give a clear articulation of an author's point of view;
- identify the most reasonable statement of an author's point of view;
- recognize bias, narrowness, and contradictions in the point of view behind an excerpt;
- identify assumptions and implications of a writer's point of view;
- distinguish evidence from conclusions based on that evidence;
- give evidence to back up their position in an essay;
- recognize data that would support, data that would oppose, and data that would be neutral with respect to, an author's position;
- recognize conclusions that go beyond the evidence;
- note, in an evaluative essay, the presence, or the absence, of evidence in an excerpt;
- identify the main concepts in a passage;
- distinguish central from peripheral concepts;
- identify the assumption underlying a given inference;
- evaluate the aptness of different versions of an assumption;
- choose the most reasonable statement of a background theory involved in a passage;
- distinguish between inferences and assumptions;
- rank different formulations of assumptions with respect to which is the most reasonable;
- identify crucial implications of a passage;
- discriminate between consequences that are necessary, probable, and improbable;
- evaluate an author's inferences;
- make, in an evaluative essay, justified inferences;
- choose the most accurate version of an author's inferences;
- draw reasonable inferences from positions they disagree with.

b) Abilities

Abilities, grounded in a thorough familiarity with the elements of thought, are the activities we actually use to perform our higher order thinking. Abilities like clarifying values and standards, comparing analogous situations, generating and assessing solutions, analyzing and evaluating actions or policies are the stuff of reasoning. They are the means whereby decisions are to be made, problems are to be solved, thinking in the work-place is to be strengthened, and understanding of rights and responsibilities deepened.

The abilities of critical reading and critical writing are keystones of any process to assess higher order thinking in that each of them, when considered at any level, is permeated by other critical thinking abilities. It is not as if we read *and* clarify values, read *and* compare analogous situations, write *and* generate solutions. To read critically *is* to clarify values, compare analogous situations, and to exercise the other abilities as well; to write *is* to generate solutions and much more besides.

Assessment of proficiency in the abilities can be keyed to student performance on test items that are geared to as many of the abilities listed on p. 101 as is feasible given the time constraints of the test.

c) Affective Traits

Without assessing affective traits, only a diminished idea of critical thinking will be addressed.

What allows us to confront our prejudices and analytically break them down is not just abilities but a *commitment* to use them for this purpose. What allows us to solve our problems in a sufficiently diligent way as to address complicated and intricate real-life problems, is again not just cognitive abilities. It is intellectual perseverance — a drive, a disposition, an affective trait. A similar point can be made for each of the intellectual traits which are the driving force behind sound and penetrating reasoning.

Assessment of the affective dimensions will concentrate on those aspects it is plausible to test for within the constraints imposed by a national assessment. These will include aspects of fairmindedness, of the willingness to suspend judgment, of intellectual courage and intellectual integrity.

d) Intellectual Standards

Assessment has to involve explicit universal standards. If we are not testing students' abilities to be relevant, precise, logical, consistent, and the rest, then we are not assessing students' abilities to engage in higher order thinking. And if testing personnel do not employ these same explicit standards, then they are grading for something other than higher order thinking.

Relative mastery of these intellectual standards requires students to be able to

- recognize *clarity* vs. unclarity;
- distinguish *accurate* from inaccurate accounts;
- decide when a statement is *relevant* or irrelevant to a given point;
- identify inconsistent positions as well as (relatively) *consistent* ones;
- discriminate *deep, complete, and significant* accounts from those that are superficial, fragmentary, and trivial;
- evaluate responses with respect to their *fairness;*
- prefer *well-evidenced* accounts to accounts that are unsupported by evidence;
- tell *good reasons* from bad.

2) Varieties of Assessment Strategies

The assessment should contain three kinds of items: *A)* machine-gradable multiple-choice items; *B)* machine-gradable multiple-rating items; C) essay items.

A) Multiple-Choice Items

Legitimate use of multiple-choice items on the assessment is limited. This type of item is geared toward relatively straightforward skills of reasoning, particularly with respect to recognizing elements of thought, distinguishing one element of thought from another, and recognizing clear examples of faulty reasoning.

Two detailed samples of assessment items follow (the first, Figure 3, is on Inferences, the second, Figure 4, on Recognition of Assumptions.)

Other abbreviated samples of appropriate multiple-choice items are as follows:

1) In the following excerpt, mark E for each item that is a piece of empirical *evidence;* mark C for each item that is a *conclusion* based on evidence; mark N for each item that is neither....

2) In this test, each exercise consists of several statements (premises) followed by several suggested conclusions.... If you think the conclusion *necessarily* follows from the statements given, make a heavy black mark under *"Conclusion Follows"*; if you think it is not a necessary conclusion, put a mark under *"Conclusion Does Not Follow."*

3) The following is a list of possible findings in relation to the experiment quoted above. For each, say whether it would *support* the author's hypothesis, *oppose* the author's hypothesis, or be neutral with respect to the author's hypothesis....

4) Below is a series of questions. Each question is followed by several reasons. For the purpose of this test, you are to regard each reason as true. The problem then is to decide whether it is a *strong reason* or a *weak reason*....

5) Which of the following conclusions is C completely supported by the stated evidence, *P* partially supported by the stated evidence, or *U* unsupported by the stated evidence?

6) Which of the following is an *implication* of the author's position in the passage cited?

B) Multiple-Rating Items

Though the use of multiple-choice questions is justified in assessing some micro-skills, the bulk of the machine-gradable items will be *multiple-rating* rather than multiple-choice. Multiple-rating items allow one to ask questions where any number of answers from a provided list may be correct, or incorrect. It further allows students to *rank*, from a number of possibilities provided, those that are

Inferences

DIRECTIONS: An inference is a conclusion a person can draw from certain observed or supposed facts. For example, if the lights are on in a house and music can be heard coming from the house, a person might infer that someone is at home. But this inference may or may not be correct. Possibly the people in the house did not turn off the lights and the radio when they left the house.

In this test, each exercise begins with a statement of facts that you are to regard as true. After each statement of facts you will find several possible inferences — that is, conclusions that some persons might draw from the stated facts. Examine each inference separately and make a decision as to its degree of truth or falsity.

For each inference you will find spaces on the answer sheet labeled J, PJ, ID, PU, and U. For each inference make a mark on the answer sheet under the appropriate heading as follows:

J if you think the inference is definitely JUSTIFIED; that it properly follows beyond a reasonable doubt from the statement of facts given.

PJ if you think the inference is PROBABLY JUSTIFIED; that it is more likely to be true than false in the light of the facts given.

ID if you decide that there are INSUFFICIENT DATA; that you cannot tell from the facts given whether the inference is justified or not; if the facts provide no basis for judging one way or the other.

PU if you think the inference is PROBABLY UNJUSTIFIED; that it is more likely to be false than true in the light of the facts given.

U if you think the inference is definitely UNJUSTIFIED; that it does not follow, either because it misinterprets the facts given, or because it contradicts the facts or necessary inferences from those facts.

Example

The first newspaper in America, edited by Ben Harris, appeared in Boston on September 25, 1690, and was banned the same day by Governor Simon Bradstreet. The editor's subsequent long fight to continue to publish his paper and print what he wished marks an important episode in the continuing struggle to maintain a free press.

1) The editor of the first American newspaper died within a few days after his paper was banned on September 25,1690.

2) Information about the first issue of Ben Harris's newspaper promptly came to Governor Bradstreet's attention.

3) The editor of this paper wrote articles criticizing Governor Bradstreet.

4) Ben Harris persisted in holding to some of his aims.

5) Governor Bradstreet objected to some of the items published in Ben Harris's paper.

In the above example:

Inference 1 is (U) unjustified because in the facts given it mentions "the editor's long fight to continue to publish his paper..."

Inference 2 is (J) justified because the facts state that the first newspaper appeared on September 25, 1690, and was banned the same day by the Governor.

Regarding inference 3, there is no information given about the precise nature of the articles appearing in the paper; thus (ID) Insufficient data.

Regarding inference 4, the facts given mention "the editor's subsequent long fight to continue to publish his newspaper and print what he wished..."; thus (J) justified.

Inference 5 is deemed (PJ) probably justified because the Governor banned the paper the day it appeared. However this is PJ rather that J because there may have been reasons for the ban other than objections to some of the items that appeared in the paper.

figure 3

more correct. Thus students can be tested on their ability to arrange items on a continuum of reasonability. This allows much more subtle testing and grading.

The same list of possible answers can pertain to any number of independent test items. Thus, a list of twenty possibilities can be provided, and students can be asked to choose the appropriate response from that list to six different questions. There is no restriction on the number of times a given answer may be correct. Nor is there any guarantee that there will be a reasonable answer on the list to every question. Guessing, using the process of elimination, and scoring well because of test-taking skills are all but impossible.

By including clearly unreasonable choices among the multiple-rating possibilities, a grade can be much more sensitive to the *degree* of a macro-ability or to the *intensity* of an affective dimension. Thus, if there are five possible answers to a given question, they need not be graded 5, 4, 3, 2, 1. Rather, they may be graded, say, 5, 4, 1, 1, -3.

We have provided two detailed samples of multiple-rating items: Figure 5 is on Reasoning Within Conflicting Points of View (and thus is an assessment of an aspect of the affective trait of fairmindedness) and Figure 6 is on Comparing Analogous Situations (and is thus an assessment of a macro-ability). Each sample is limited here by having only four possible answers, a limitation that would not obtain on an actual test.

The following is a list of abbreviated samples of multiple-rating items, having to do with elements of thought, with abilities, with affective dimensions, and with intellectual standards.

Multiple-Rating Items, Elements of Thought

- Here is a list of formulations of the writer's objectives in this excerpt. Rank them from 1 to 5 with respect to which is the most reasonable in the light of the quoted passage....
- For each of the underlined passages in the excerpts below, mark P on the answer sheet if it is a statement of the writer's *Purpose*, *C* if it is a statement of the *Consequences*, *A* if it is a statement of the writer's *Assumptions*, and *I* if it is an *Inference* the writer is making.
- Which of the following would the author most likely give as the statement of the problem she is attempting to solve?
- Read the excerpt; then, from the following list, identify the most plausible statement of the writer's purpose.

Recognition of Assumptions

DIRECTIONS: Careful reasoners often find it necessary to complete partially stated arguments in order to evaluate those arguments. For example, someone might say, "John is selfish; we are good friends, but he never lends me money." The conclusion that "John is selfish" is supported by two explicit claims:

1) John never lends me money.
2) John and I are good friends.

But an important part of the argument was left out:

3) People who never lend money to their good friends are selfish.

This third assertion is an *unstated assumption* of the argument.

In this test each exercise begins with a brief argument. Each argument is followed by three numbered statements. Examine each of the numbered statements individually and make a decision about its logical relationship to the argument. For each numbered statement there are spaces on your answer sheet labeled: EC, UA, and N. Select just one of the following alternatives for each numbered statement, and make a mark on your answer sheet under the appropriate heading:

EC if you think the idea expressed in the numbered statement is an *explicit claim* made in the argument (even if the wording is not the same).

UA if you think the idea expressed in the numbered statement is a probable *unstated assumption* of the argument.

N if you think the idea expressed in the numbered statement is *neither* an explicit claim nor an unstated assumption of the argument.

Example

Argument: "We need to save time in getting there, so we'd better go by plane."

1) Going by plane will take less time than going by some other means of transportation.

[Saving time is given as a reason for going by plane; this only makes sense if the person giving the argument believes that going by plane would take less time than other available means of transportation. So the idea expressed here is an *unstated assumption* of the quoted argument.] (UA)

2) We should try to cut down how long we spend travelling to our destination.

[The idea expressed here is directly asserted, though in different words, in the argument, so it is not an unstated assumptions of the argument; rather, it is an *explicit claim* made in the argument.] (EC)

3) Travel by plane is more convenient than travel by train.

[No mention is made in the argument of either trains or convenience. The idea expressed here is *neither* an explicit claim nor an unstated assumption of the argument.] (N)

figure 4

- Of the following statements of the author's *point of view* in this passage, select the one from the following list that is both most reasonable and most relevant to the passage....
- List A below is a list of various possible statements of the writer's point of view in the quoted passage; List B is a list that includes possible assumptions and implications of those points of view. Match the items on list A with the items on list B...
- Which of the following are main *concepts* in the passage cited; which are *peripheral concepts*?
- For each inference below, decide whether the accompanying statement is *U* an unstated assumption, *A* an assertion, or *N* neither...
- Rank the following items on a scale of 1 to 5 according to how reasonable it is as a statement of the author's *assumptions*...

Reasoning Within Conflicting Points of View

Directions: In the following questions, rank the answers in order of reasonability. In each case you are being asked to rank answers as to which is the strongest argument in favor of a position. By the strongest we mean the one that is most defensible, not necessarily the one which claims the most. To rank a defense for a position high does not mean that you actually hold that position but only that if you had to defend it before an audience of unbiased and openminded people, the options you rank higher would be easier to defend on rational grounds than the ones you rank lower.

1) Children under the age of twelve should have all of their important decisions made for them by their parents and other appropriate adults because:

1) allowing them to make all important decisions for themselves will encourage false pride and stubbornness.
2) allowing them to make all important decisions for themselves will undermine parental respect and authority.
3) children are not mature enough to make all important decisions for themselves.
4) children should not be expected to take life's problems so seriously until they grow up.
5) children can be expected to make grave mistakes, some of which could harm them for life.

2) Children under the age of twelve should make some important decisions for themselves because:

1) children are less prejudiced than adults and more open to the truth.
2) children spend a lot of time watching T.V. so they know a lot about what is going on in the world.
3) children are likely to make many reasonable decisions affecting themselves.
4) children will become depressed if they are not allowed to make some important decisions.
5) children will be more apt to become responsible adults if they are allowed to make some important decisions for themselves as they are growing up.

BEST COPY AVAILABLE

Comparing Analogous Situations

"Having a population to study instead of an individual fossil is enormously important. No two people today are exactly alike; no two Australopithecines were either. It is for that reason that drawing conclusions from a single fossil is risky. Measurements taken of it, and theories spun off as a result of those measurements, may be misleading because the part being measured may not be typical. It is only when a large number of specimens is available that all their variations can be taken into account, and a norm derived from them. If a visitor from outer space were to describe and name *Homo sapiens sapiens* by examining one skeleton, that of a short, squat, heavy-boned New Guinea tribesman, he would certainly be excused if he set up another species on the basis of a second skeleton discovered later a few thousand miles away — that of a seven-foot, slender-boned Watutsi tribesman from central Africa." (Edey, *The Emergence of Man*, pp. 47–48)

The author of the above passage makes an analogy between an anthropologist studying fossils and a visitor from outer space studying one or two single skeletons. Rank each of the following comments 1 to 3, according to whether it would be crucial to judging the strength of the analogy for the point the author is making. Give a comment a 3 if it is CRUCIAL in judging the worth of the analogy; give it a 1 if it is IRRELEVANT to judging the worth of the analogy; give it a 2 if it lies in between.

a) The analogy illustrates the point well because in both cases we are called upon to draw general conclusions based on a limited sample. The more items you have in your sample, the more justified your generalization will be.

b) It is a bad analogy because the visitors from outer space would draw the same erroneous conclusion even if they had a whole population of New Guinea tribesmen to study.

c) It is a good analogy but it shows that we need, not simply *more* fossils of Australopithecus, but fossils of it from other geographical areas.

d) It is a bad analogy because we have no idea what visitors from outer space would conclude from seeing a skeleton of a New Guinea tribesman. The visitors might refrain from making the generalization for the same reason that makes the author say it is "risky."

figure 6

- Look at each of the statements below as a possible consequence of the writer's position in the excerpt cited. Rank each statement on a scale of 1 to 7, where 7 means that you consider the statement a *necessary* consequence of the passage, and 1 means that you consider the statement a *highly unlikely* consequence of the passage.
- Each of the following is an *inference* one might draw from the passage. Rank each one on a scale from 1 to 5, according to whether it is completely justified (5) or completely unjustified (1)...
- Which of the following is the most accurate formulation of the author's *inference* in the cited passage?

Multiple-Rating Items, Abilities

- Which of the following would be relevant to deciding whether A is a *credible* source of information on the topic...?
- Here is a list of observations about the behavior of X's, made by a responsible investigator. Which of the items from the following list would be a *justified generalization* about X's?
- A has the following beliefs about astrology. Which of the questions below would be *root* or *significant questions* that A would have to answer to claim his beliefs about astrology were rational?
- A refuses to refund a customer's money and, when asked, defends her action by stating that it is "dictated by store policy". Which of the following would be *relevant* to deciding whether her action was indeed "dictated by store policy"? Which of the questions would be relevant to deciding if the store policy was rational?
- Judge A makes the following ruling in a case... Which of the following is the clearest statement of the standards Judge A is using?
- A compares the relation between managers and employees to the relation between teachers and students. Which of the following would A have to answer in order to continue using the analogy rationally?
- A gives the following argument for.... Which of the listed comments would be the *strongest objection* to her argument?
- Listen to the accompanying excerpt from an audiotape of a lecture by A. Which of the following questions would be of most help in clarifying A's views?

Multiple-Rating Items, Affective Traits

- Here are position-statements from both sides, A and B, of a controversial and inflammatory debate. From list X below, choose those items which are the most reasonable *inferences* to draw from position A; then choose those items which are the most reasonable *inferences* to draw from position B.
- Here are position-statements from both sides, A and B, of a controversial and inflammatory debate. From list X below, choose those items which state the most reasonable *assumptions* underlying position A; then choose those items which state the most reasonable *assumptions* underlying position B.
- For each of the items below, tell which is the most reasonable action to take under the circumstances described. If, in your view, there is not enough information to make a reasonable decision, you may choose the action of *suspending judgment* as the most reasonable response.

BEST COPY AVAILABLE

- A disposition to take a measured response rather than an exaggerated, disproportionate response will be measured by requiring students to discriminate between the likelihood of dire versus mild consequences of positions they dislike.

Multiple-Rating Items, Intellectual Standards

- The following are four definitions from *Webster's New World Dictionary*. Which of them gives the clearest definition of...?
- Rank the following definitions for their *precision* on a scale of 1 to 7. 1 means "not precise at all"; 7 means "too precise for the subject matter"; and 4 means "exactly as precise as it should be".
- Here is a list of data and a series of accounts summarizing the data. Which of the accounts is the *most accurate* summary of the data?
- For each statement below, tell whether it is *relevant* or *irrelevant* to the hypothesis in the passage cited.
- Which of the following is the *fairest* restatement of the author's position [where the author is stating a highly controversial position]?
- Rank the following statements according to which are the *best-evidenced* and which are the *least-evidenced*.
- Which of the following is a good reason for believing the statement in question? Which is a bad reason? Which is somewhere in the middle?

c) Essay Items

The full range of the use of critical thinking cannot be assessed without requiring writing on the part of the student. To confront real issues, balance competing interests, weigh objections and alternatives, and make a reasonable decision about a matter of some consequence — this is a major part of what it is to think critically.

The ability and the disposition to engage in full-fledged critical thinking is measured only in part by a person's ability to choose from among a pre-selected list. A true measure of critical thinking, and thus of a program's capacity to improve critical thinking, can be obtained only by including in the assessment *generative* as well as *selective* dimensions. Neither multiple-rating nor, obviously, multiple-choice items are adequate for testing this dimension.

Essay items will require proficiency in handling the elements of thought, in using appropriate abilities, in applying intellectual standards, and, what is more, it will require integrating these and bringing them to bear on a substantive issue.

Three detailed samples of essay items follow on the next page. Each has the same set of general directions.

In addition to full-blown essay tests, a series of short-justification items are currently being prepared. These would not ask students to write an essay on a topic, but would rather have them choose an answer from a pre-selected multiple-rating list and then justify their answer in a sentence of their own writing.

This type of test, if it were sufficiently developed, would have several advantages: it could be administered, because of the brevity and straightforwardness of students' written answers, to the student population as a whole rather than merely to a representative sample (see #1, under "Implementation", below); it would assess some, though not all, generative dimensions of critical thinking; it would allow flexibility in grading the machine-gradable keyed answers (thus, one could adjust the rating of an item up or down depending on the justification); it would be no more difficult to grade by trained personnel than the math work on currently administered standardized calculus tests.

✦ *Interdisciplinary and Subject-Specific*

Scope of the Assessment

An assessment of the results of critical thinking instruction ought to focus both on thinking within the framework of particular academic subjects, and on thinking in the interdisciplinary contexts that are so important to functioning as an autonomous, well-informed, productive member of a democracy.

A basic principle of critical thinking instruction, as applied to teaching subject matter in an area, is that (to quote the National Council for Excellence in Critical Thinking Instruction) "to achieve knowledge in any domain, it is essential to think critically". A related principle is that in any domain where one is thinking well, one is thinking critically. Any example of good scientific thinking, or good historical thinking, or good anthropological thinking, or thinking in any other subject, will necessarily be an example of critical thinking: It will involve basic skills dealing with elements of thought; it will involve at least some, and probably many, of the abilities; it will involve affective traits like independent thinking and intellectual perseverance. And as far as instruction is concerned, there is a real sense in which learning biology is learning to think within and about the logic of biology.

Including critical thinking items taken from individual subject areas would also properly test those thinking skills that are more subject-specific, and it would do so in the context of presupposing a good deal of specialized knowledge. A critical thinking test in nursing or in history of art or in geology might well (in their different ways) test for skills of critical observation, while a test in sociology might assess thinking skills involved in constructing an unbiased questionnaire; a critical thinking test in English literature might well presuppose a knowledge of who Milton was, while a thinking test in physics might justifiably ask about a problem for which a knowledge of the second law of thermodynamics was taken for granted.

Critical Thinking, Problem Solving, & Communication Skills Essay Exam

Directions

This test is designed to assess your critical thinking, problem solving, and communication skills. Your answer will be judged for its clarity, relevance, consistency, logic, depth, coherence, and fairness. More specifically, the reader will be asking the following questions:

1) Is the question at issue well stated? Is it clear and unbiased? Does the expression of the question do justice to the complexity of the matter at issue?
2) Does the writer cite relevant evidence, experiences, and/or relevant information essential to the issue?
3) Does the writer clarify key concepts when necessary?
4) Does the writer show a sensitivity to what he or she is assuming or taking for granted (insofar as those assumptions might reasonably be questioned)?
5) Does the writer develop a definite line of reasoning, explaining well how he or she is arriving at his or her conclusions?
6) Is the writer's reasoning well-supported?
7) Does the writer show a sensitivity to alternative points of view or lines of reasoning? Does he or she consider and respond to objections framed from other points of view?
8) Does the writer show a sensitivity to the implications and/or consequences of the position he or she has taken?

ISSUE #1: ECOLOGY

The nation is facing a variety of ecological problems that have the following general form: an established practice, whether on the part of business and industry or on the part of the public, is contributing to serious health problems for a large number of people. At the same time it would be costly to modify the practice so as to reduce the health problem.People often say that the answer is one of achieving a "balance" between the amount of money we spend to correct the problem and the number of lives we would save by that expenditure. Develop a point of view and some plausible criteria for telling how one would determine this "balance." Make sure you address any dilemmas inherent in your strategy for solving such problems.

ISSUE #2: POLITICS

There is a growing number of Americans who do not vote in national and local elections. Many of them explain their non-participation by saying that their vote would not make a difference. Some go on to argue that this is true because "money plays such a large role in elections that the candidate with the highest paid, and the highest quality, media campaign wins." Most people agree that money sometimes plays an inappropriate role in determining the outcome of elections. Develop a proposed solution to this problem that takes into account the view that people and organizations with money have a right to use that money to advance political causes they believe in. If you like, you may decide to develop a position to the effect that there is no solution to the problem and that we have no choice but to accept the status quo.

ISSUE #3: MORALITY

Sociologist Erving Goffman has pointed out that all social groups, including professions, develop a protective attitude toward members of their group, even when what some of the members do is seen as morally wrong. A sense of loyalty to the group often overrides what they would otherwise deem immoral. Consider the arguments for and against exposing people with whom you are personally close or with whom you have close professional ties. Develop a position on this issue that could serve as a guide for anyone in such a position.

Even if we already had a series of critical thinking items within the various subject areas, however, we would not be testing for many of the interdisciplinary abilities we most want critical thinking for. Many of these have already been mentioned: the ability to make sound decisions in the context of understanding our rights and responsibilities as citizens, in the context of the work-place, as well-informed and thinking consumers, as members of our families, as participants in what is becoming a symbiotic and fragile world economy — the ability to reason about the gaps between subject areas, the bridges between them, and the generalizability of subjects to other areas.

To test critical thinking abilities, as they apply to these areas, what is needed are interdisciplinary questions. These are questions of broad interest, ones that shed light on the quality of and improvement in student thinking about realistic and fundamental issues; they ought to be the kind of questions which can be at least partially illuminated by well-integrated knowledge in any number of academic areas.

The national assessment we are proposing would offer a range of subject-specific items, from which students would choose those relevant to their subject-matter knowledge. The interdisciplinary items, on the other hand, would not provide choices because of the desirability of avoiding the loss of equivalency that is almost always involved. (That loss would have to be minimized in the case of subject-specific items by field testing and rewriting.)

The interdisciplinary part is constructable by experts well versed in a rich and substantive concept of critical thinking. Subject-specific critical thinking assessment items will be constructed by members of the discipline working in consultation with experts in critical thinking, perhaps the standing committees on the various disciplines of the National Council for Excellence in Critical Thinking Instruction. Both groups would work in conjunction with grade-level experts to construct appropriate levels of items, from the 6th-grade test through the college-graduate test.

✦ *The Value of the Proposed Assessment Strategy for the Reform of Instruction*

Since higher order thinking has always been considered an important object of education, and since this assessment would furnish a measure of that concept, and since performance on this assessment would have a significant impact on the standing of the school not only in the eyes of the intellectual community but in the eyes of the public as well, administrators and teachers would have a strong motivation to become familiar with the concepts and program behind the assessment. Most importantly, teachers and others in charge of instruction and the formulation of educational goals would find in it a clear model for the articulation and integration of higher order thinking across the curriculum. Note the following:

1) The concept of the elements of thought not only provides a realistic analysis of the common dimensions of reasoning in every domain, it also encourages the explicit use in instruction of those critical/analytic terms which are the common possession of the intellectual community (question-at-issue, problem, evidence, data, concept, inference, assumption, implication, conclusion, point of view, frame of reference, etc.) and makes explicit the intellectual standards implicit in every subject as well as in the closely reasoned professional work in business and industry (clarity, precision, accuracy, logic, consistency, ...)

2) By highlighting reading, writing, speaking, and listening as modes of critical reasoning, the necessity of having instruction go beyond mere didactic coverage of content would become more intelligible. As long as reading, writing, speaking, and listening skills appear the sole province of specialized subjects and at specialized levels rather than modes of reasoning intrinsic to the construction and mastery of knowledge in any subject at any level, there will continue to be a significant lack of fit between modes of instruction and modes of necessary learning.

3) By highlighting the other abilities of critical thinking, each analyzed into the same elements of thought, there would be significant transfer of emphasis to important modes of higher order thinking within a larger number of student assignments. At present, many teachers fail to notice the extent to which they either presuppose that students already grasp the nature of fundamental intellectual processes, or they make assignments which, though they appear to call for such processes, can be successfully completed by simply repeating to the teacher what was said in lecture or written in the text.

4) By highlighting a common critical/analytic language across the curriculum, students are encouraged to seek to transfer learning and intellectual discipline emphasized in one domain of learning to other domains of learning and application. The fragmentation of the subject areas, in the minds of the students if not in fact, is now a serious problem in education. This problem is mirrored, of course, in business, industry, and government in the tendency to engage in fragmented, over-specialized problem-solving which fails to address the multi-dimensional nature of many complex problems.

5) By highlighting the importance of intellectual discipline and grounding it in specific skills and abilities, teachers and other educational leaders will be given a reasonable impetus to help students make connections of a broader, more interdisciplinary nature. This will also be strongly re-enforced by the inclusion of everyday, multi-logical, interdisciplinary essay questions.

✦ *Implementation of the Proposed Assessment*

Our recommendations about implementation can be summarized as follows:

1) The essay assessment should be administered to a representative sample of the student population at each educational institution, the machine-gradable items to the total student population;

2) it should be administered at the 6th, 9th, and 12th grades, and three times during a student's college career — at entrance, at the start of the junior year, and just prior to graduation — and thus yield value-added information to schools;

3) the test should be constructed to be roughly three-hours long;

4) test items should be constructed from item shells, rather than from a simple pool of actual items;

5) it should be administered by a private agency with critical thinking credentials;

6) it should be paid for by school districts, colleges, and universities that contract to have their students tested;

7) it should provide educational institutions with detailed information about central aspects of their students' higher order thinking;

8) it should be developed according to the costs and timetables listed below.

Details of our recommendations center around the answers to five practical questions about the administration of the test:

Who Will Be Assessed?

Our *minimal* recommendation is that all portions of the assessment be given to, at the very least, a representative sample of the student population at each educational institution. Since the problems implicit in testing a random sample can be easily worked out, this recommendation avoids the expense of administering an essay test to the student population as a whole.

The assessment strategies we have proposed include two broad areas of testing: a *machine-gradable portion* that includes multiple-choice items and multiple-rating items and an *essay portion*. Both portions will assess, in their different ways and with their different emphases, micro-skills, abilities, affective traits, and intellectual standards.

There are, therefore, really two options with respect to who is assessed using the strategies we propose. First, the machine gradable portion of the assessment can be administered to the student population as a whole, while the essay portion can be administered to a representative sample of students at each institution. Second, both portions could be given only to a representative sample of the population at each institution. Both options will hold down costs, though the latter will clearly be less expensive than the former. Which option is ultimately chosen will depend on the amount of detail desired, the precise role the assessment is to play, and the funds available.

How Often Will the Assessment Take Place?

The maximum benefit to educational institutions will be provided to the extent that they are enabled to measure the progress of their students' higher order thinking during the course of their educational career. This will enable school systems not only to gauge

their contribution to their students' progress, but also to measure the success of attempts to re-design their instruction so as to increase critical thinking capabilities.

These objectives can be accomplished by having students assessed often enough to reflect such progress, optimally: at the 6*, 9*, and 12* grades, and at the time of their college entrance, at the beginning of their junior year, and just before graduation from college.

How Long Will the Test Take?

The test should last about three hours in order to cover multiple-choice, multiple-rating, and essay items without becoming a speeded test to an inappropriate degree. To span all difficulty levels, it would be best to have a total of at least 30 items. While two of these could be short essay items requiring 20 minutes each to answer, the machine-gradable items would be faster to answer, and hence could be handled in 3–8 minutes.

How Will a Sufficiently Large Pool of Items Be Constructed?

While it might be possible to release a pool of items which would provide the equivalent of 6 tests at each level, hence 6 x 6 x 30, it would be better to increase flexibility by using item shells, which would be items that include identified variables, each of which could be replaced from a list of acceptable values. This would greatly increase the number of items that could be generated, but without "surprises". A pool of shells would generate over a thousand items at each level, possibly several thousand.

Who Will Do the Assessing?

In order to avoid problems in the reliability of the assessment (like those we have seen occur in the *California Direct Writing Assessment*), the assessment needs to be monitored, administered, and graded by a private agency whose personnel have critical thinking credentials or are at least under the direction of scholars with a solid grounding in research in critical thinking.

Who Will Bear the Costs of the Assessment?

The assessment should be paid for by the school systems, colleges, and universities that contract to have their students tested. This not only puts least burden on the public but represents an established precedent in distributing costs of testing.

What Will Institutions Be Able to Learn from the Results of the Assessment?

We anticipate that educational institutions will receive an analytic report that will document all of the following:

- where their students are strongest and weakest with respect to particular micro-skills;
- where their students are strongest and weakest with respect to important abilities;
- how students stand in each of the school's subject-matter areas;
- how their students stand in relation to students at other institutions;
- how their students at one educational level stand in relation to their students at other educational levels;
- how their students stand with respect to established performance criteria.

This information would enable institutions to target instruction to remediate weaknesses and build on strengths, as well as to measure what students are gaining as a result of attending their classes.*

✦ *Postscript*

Dale Carlson, the head of the California Assessment Program, has recognized that the essay "Rock Around the Clock" was significantly mis-assessed and is now devising strategies to prevent this from happening in the future.

* The authors wish to acknowledge the invaluable advice provided us by Michael Scriven on evaluation theory in general, and, more particularly, on the logistics of test construction.